

SZOTE Központi Laboratórium Számítástechnikai Központ

Adattfeldolgozás céljait szolgáló programrendszer R-10-es számítógépre, különös tekintettel biometriai feldolgozásra

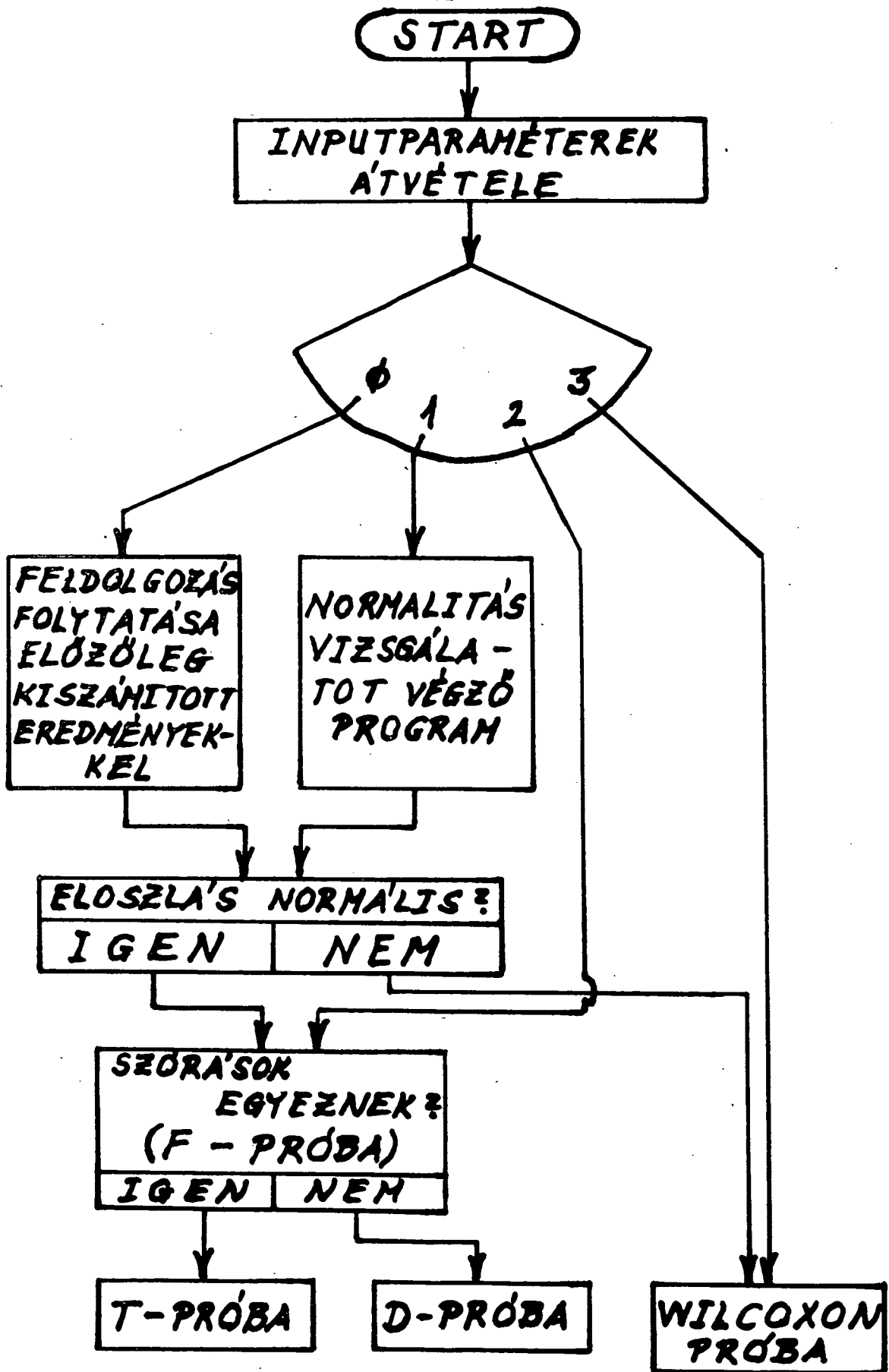
Szeberényi László, Kasza Ferenc, Stéhlik Jánosné, Győri István és
Forczek Erzsébet

A SZOTE Központi Laboratórium Számítástechnikai Központ feladatainak jelentős részét - több éves tapasztalat alapján - az egyetem egyes intézeteitől kapott biológiai jellegű adatok statisztikai feldolgozása képezi. Az ezirányban felmerült számítástechnikai igények kielégítése eddig nagyrészt egyedi, egymástól részben, vagy teljesen független programokkal történt. A feladatok számának növekedése és a számítógépes szemlélet elterjedése révén a problémák új oldalról való megközelítése szükségessé tette egy hatékonyabb programrendszer kiépítését.

Ennek a programrendszernek igen összetett, egymásnak sokszor ellentmondó igényeknek kell eleget tennie. Hogy ezeket az igényeket és a rendszertervezés kezdeti stádiumában felmerült problémákat szemléltetni tudjuk, nézzük meg egy gyakran előforduló, mindenki számára ismert problémakört (1. ábra): két független minta várható értékének összehasonlítását. Itt a leghatásosabb próba a 2 mintás t -próba. De mint tudjuk, ez csak akkor végezhető el, ha a két minta normális eloszlású populációból származik, és a minták szórásai szignifikánsan nem különböznek.

Ha nem normális eloszlású populációból származó mintáról van szó, Wilcoxon próbát lehet végezni. Ha normális volt az eloszlás, meg kell vizsgálni a szórások egyezését.

Egyező szórások esetén t -próbát, egyébként D -próbát végzünk. Ahhoz, hogy egyszeri géphezfordulással készüljön a feldolgozás, a programnak magának kell tudni eldönteni, hogy melyik próbát kell alkalmaznia. Ez egyetlen program esetén a memória nagy részét lefoglalná, így a fennmaradó helyen sok adattal nem lehetne dolgozni. A feladatot részekre kell tehát bontani, több önálló rész vizsgál egy-egy feltételt. Így egy feladatot több programhívással lehetne csak végrehajtani, - az pedig nehézkessé tenné a mindennapi feldolgozást. Ezen segítene egy olyan kezelőrendszer, amely egyszerűbb fastrukturákban adott feladatokat is tud vezérelni úgy, hogy a benne



szereplő ágak önállóan is futtathatók. Ha ez teljesülne, akkor egy-egy több relációban előforduló probléma programja csak egyszer szerepelne a rendszerben, ugyanakkor elérhető - megfelelő vezérlő és futtató rendszer esetén - , hogy a feladatot leíró fastruktúra vezérlőprogramja, egyetlen ága, vagy részága legyen a tárban. Ez különösen hosszú és nagy tömbökkel dolgozó programoknál előnyös. Például az ábrán a normalitás, vagy a Wilcoxon próba, olyan statisztikai próbák, amelyeknek önállóan is létjogosultságuk van, ugyanakkor az említett példában, mint esetleges részfeladatok szerepelnek.

Előfordul, hogy egy előzetes feldolgozás folytatása esetén a normalitás kérdése már eldőlt, vagy bizonyos jellemzőket már kiszámítottunk. Ezeket újból fel tudjuk használni anélkül, hogy ismételten kiszámítanánk, ha előzőleg egységes, minden program által könnyen kezelhető formában tároltuk azokat.

Az 1. ábrán például a program 0 konstans esetén a már kiszámított eredményekkel dolgozik. Ha konstans 1-es, akkor szükség van a kérdéses mennyiségek kiszámítására. A 2-es, ill. 3-as paraméter megadja a felhasználónak azt a jogot is, hogy maga határozza meg, melyik próbát hajtja végre.

Gyakori eset, hogy a statisztikai kiértékelés előtt bizonyos mérési eredményeken transzformációt kell végezni. A rendszertől azt kívánjuk, hogy a transzformáció által létrehozott új minták ugyanugy kezelhetők legyenek a rendszer által, mint az eredetiek.

A rendszernek biztosítania kell a statisztikai táblázatok egységes kezelését is.

Ha a fenti igényeknek eleget tesz a rendszer, akkor nagy tömegű adatok egyszeri és folyamatos feldolgozása is lehetővé válik, ugyanakkor néhány adat azonnali, gyors, egyszeri géphezfordulással való kiértékelésére is alkalmas.

A fenti megfontolásokból kiindulva és figyelembe véve egy sor egyéb - most nem részletezhető - statisztikai igényeket és software-hardware lehetőségeket, a megvalósítás alatt álló feldolgozórendszer a következő alapkövetelményeknek tesz eleget:

1. A felhasználó kényelmének biztosítása végett a rendszer működtetésének programozása egyszerű, továbbá működés közben minimális emberi beavatkozást kíván.

2. A rendszerhez tartozó programok adatkezelése egységes és tömör.

3. A rendszer egyszerűen bővíthető - szűkíthető, vagy akár átalakítható azért, hogy a rendszerbe beépített biometria módszerek köre minél tágabb lehessen, valamint egymással jól variálhatók legyenek.

4. A rendszer magja, a futtatórendszer minimális állandó specifikációt tartalmaz arra nézve, hogy az egyes konkrét esetekben milyen célú programrendszer áll a vezérlése és felügyelete alatt. Ez lehetővé teszi, hogy a jelenleg készülő statisztikai rendszertől eltérő egyéb feldolgozó rendszereket alakítsunk ki a rendszer magja köré.

5. A rendszer felügyelete alatt futtatható programokkal szembeni megkötések minimálisak.

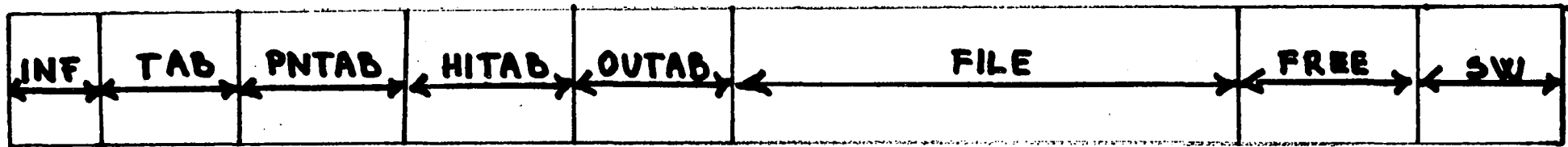
6. Komplex, több programot vagy feltételesen több programot igénylő feladatok esetén a feladat megoldási menete (a futtatandó programok kiválasztása és sorrendje) maga is programmal adható meg, ami egyben biztosítja a feldolgozás strukturájának egyszerű áttekinthetőségét.

Ugy érezzük, hogy a felsorolt követelményeknek eleget tevő rendszer lehetőséget nyújt a felhasználónak átfedésmentes programrendszer összeállítására, valamint a feladat megfelelő megfogalmazása esetén automatikus feldolgozásra. A rendszer működéséhez szükséges összes információ a disk ugynevezett "felhasználói" zónájában helyezkedik el, mely a 2. ábrán látható.

A rendszer strukturáját és egyes elemeinek funkcióit a 3. ábra szemlélteti.

Központi helyet foglal el a rendszerben a vezérlő mag, vagy más néven a futtató-rendszer, mely a mindenkori Monitorból - kiegészítve egy speciális szekcióval - valamint a HEAD rendszerprogramból áll, melynek funkciói és jellemzői a következők:

DISK "FELHASZNÁLÓI" ZÓNA



INF - az aktuális felhasználóra vonatkozó információk zónája.

TAB - tetszőleges, - a rendszer működése alatt - standard táblázatok zónája.

PNTAB - választható programok listája.

HITAB - az aktuális feladat programhívásainak listája.

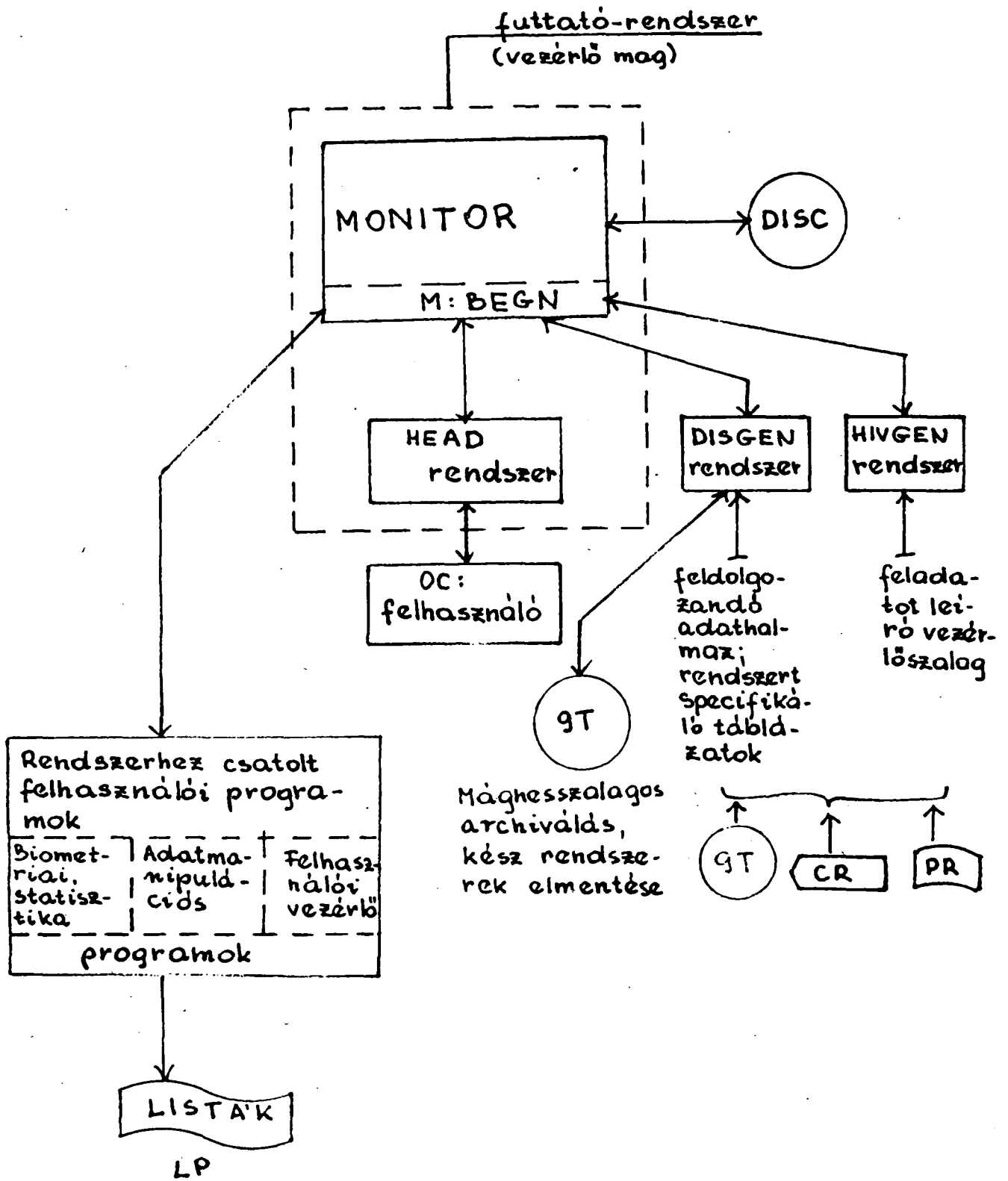
Fontos megjegyezni, hogy ez a négy táblázat specifikálja a feldolgozórendszert minden egyes működés esetén.

OUTAB - eredménytáblázatok zónája.

FILE - a feldolgozandó adathalmaz rekordjait tartalmazza

FREE - a rendszer programjainak szabad munkaterülete

SW - SWAPPING zóna, felhasználó programok időleges mentési területe.



3. ábra

- Biztosítja a rendszer és a felhasználó közti kommunikációt.
- Ellátja a feladat végrehajtásához szükséges programok indítását - futtatását a HIVGEN rendszer által a disken kialakított programhívási lista alapján.
- Az éppen futó felhasználó program részére a következő szolgáltatásokat nyújtja:
 1. Lebonyolítja a felhasználói program és a disk közötti adatforgalmat egy speciális file-kezelőrendszer segítségével, ezzel megvalósítva az egységes adatkezelést.
 2. Biztosítja felhasználó programok felhasználó programból történő indítását a hívó programhoz történő visszatéréssel vagy visszatérés nélkül, ezzel lehetőséget adva a felhasználónak vezérlő-programok kialakítására.

Nagyon fontos, hogy a vezérlő mag a file-kezelőrendszeren kívül semmiféle állandó specifikációt sem tartalmaz a vezérelt programrendszerre nézve.

A rendszer specifikálását a DISGEN rendszerprogram végzi el, kialakítva az egyes egyedi rendszereket, továbbá lehetővé teszi az eredmények, illetve adott esetben az egész rendszer archiválását. A feldolgozandó adathalmaz diszkre juttatását szintén a DISGEN végzi el.

Az ily módon kialakított rendszernek az elvégzendő feladatot, a futtatandó programok listáját, az aktuális paramétereikkel együtt, a HIVGEN rendszer szolgáltatja a feladatot leíró vezérlőszalag vagy vezérlőkártyák alapján. A feladatok leírására egy egyszerű célnyelvet fejlesztettünk ki, melynek feldolgozását az említett program végzi el.

A rendszer hardware igénye 32 Kbyte belső memória, minidiszk vagy 5 Mbyte-os cserélhető diszk, sornyomtató, lyukszalag vagy kártyaolvasó, esetleg 1 mágnesszalagos egység.

1977. áprilisára a rendszerprogramok, valamint a legalapvetőbb statisztikai eljárások programjai elkészülnek. A rendszer könnyen bővíthető, tekintve, hogy a rendszerhez olyan programnyelveken írt programok csatolhatók, amelyekben az ASS2 és a másik programnyelv közötti információ csere megoldott. Így teljes kiépítés után lehetőség nyílik a legkülönbözőbb szignifikancia vizsgálatok elvégzésére, a variancia analízis, a

faktoranalízis, az orvosi diagnosztikai célokat szolgáló programok rendszerben történő futtatására.

Az elhangzott előadás anyaga az EU.M. 4-13-0201-03-0/GY számú "Számítástechnikai módszerek, rendszerek, berendezések fejlesztése adaptálása az orvostudományban és az egészségügyben" c. tárcaszintű kutatási főirányhoz minisztériumi szinten kiemelten elfogadott "Számítástechnikai és matematikai módszerek alkalmazása az orvostudományban és az egészségügyben" c. témában végzett kutatómunka alapján készült.